

# Best Google Indexing Site for Website Crawl Status Analysis

[ExpressIndexing](#) provides website owners, SEO professionals, agencies, publishers, bloggers, developers, and link builders with an organized way to submit URLs and monitor their progress. Its bulk submission capability, individual URL statuses, live batch tracking, exportable records, and REST API access can support a structured website crawl status analysis workflow.

Publishing a webpage does not automatically make it visible in Google Search. Google must discover the URL, request it from the server, crawl and render its content, interpret technical instructions, and decide whether the page should be included in the search index.

Website crawl status analysis helps teams understand whether search engine crawlers can access important pages efficiently. It can identify server errors, crawl restrictions, redirect problems, orphan pages, outdated sitemaps, and other conditions that may prevent valuable URLs from receiving proper search consideration.

By combining technical crawl analysis with organized URL submission and monitoring, website owners can create a clearer and more scalable discovery process.

## What Is Website Crawl Status Analysis?

Website crawl status analysis is the process of examining how search engine crawlers discover and access website URLs.

It may involve reviewing:

- Newly published URLs
- Submitted pages
- Individual submission statuses
- Googlebot requests
- HTTP response codes
- Robots.txt rules
- Redirect chains
- XML sitemap entries
- Internal links
- Crawl depth
- Server response times
- JavaScript rendering
- Mobile crawler access
- Duplicate URL patterns
- Historical crawl activity

Crawling is different from indexing. A page can be crawled without being included in Google's index. Google may access a page and later exclude it because of a noindex directive, duplicate content, canonical selection, limited value, or another consideration.

Crawl analysis focuses on access and discovery, while indexing analysis evaluates what happens after the page has been processed.

### **Why Website Crawl Status Matters**

Search engines need a clear and reliable path to valuable pages. If Googlebot cannot access a URL, Google has limited ability to evaluate the content for indexing and rankings.

Crawl problems may affect:

- Blog articles
- Product pages
- Category pages
- Service landing pages
- Location-specific content
- News stories
- Affiliate guides
- Business directories
- Educational resources
- Real estate listings
- Travel pages
- Client campaign content

Crawl analysis is especially important for large websites. E-commerce stores, news publishers, marketplaces, and directories may generate thousands of URLs through filters, parameters, searches, and dynamic navigation.

Without careful control, crawlers may spend time on duplicate or low-value pages while important content remains difficult to discover.

### **URL Submission and Crawl Status**

URL submission can support discovery by placing a page into an indexing workflow. However, it does not guarantee that Google will immediately crawl, index, retain, or rank that URL.

ExpressIndexing allows users to submit links and review their individual states. This helps teams confirm which URLs entered the workflow and identify unsuccessful requests.

A submitted URL may still experience crawl problems if:

- Robots.txt blocks it
- The server is unavailable

- The URL redirects repeatedly
- Authentication is required
- DNS resolution fails
- Important resources are blocked
- JavaScript does not render
- Internal links use another version
- The canonical points elsewhere
- The page returns an error

Technical validation should therefore occur before submission.

### **Bulk URL Submission for Crawl Analysis**

ExpressIndexing supports batches containing up to 200 URLs. This can help users manage larger crawl discovery campaigns.

Bulk submission may be useful for:

- New blog posts
- Product launches
- Category updates
- Service and location pages
- Website migration destinations
- Seasonal landing pages
- Guest posts
- External backlink URLs
- Client campaign links
- Pages corrected after technical review

URLs should be organized into logical groups.

An agency can create separate batches for each client. A publisher can group articles by publication date. An online store can divide links by product category.

Batch organization makes status analysis easier. If one group repeatedly fails while another progresses normally, the associated template, domain, or technical configuration can receive priority.

### **Individual URL Status Analysis**

Batch totals provide an overview, but individual URL tracking reveals precisely which links need attention.

Each submitted URL can carry its own state and reference. This makes it possible to isolate unsuccessful requests without manually reviewing an entire campaign.

Individual tracking helps teams:

- Confirm the correct URL was submitted
- Detect malformed addresses
- Identify failed requests
- Avoid duplicate submissions
- Organize technical follow-up
- Maintain campaign history
- Create client reports
- Compare results across page types

A failed submission does not necessarily mean Googlebot failed to crawl the page. It may indicate that the URL did not complete the platform's processing workflow.

Submission data should be combined with server logs, crawl reports, and technical checks for a complete understanding.

### **Understanding HTTP Crawl Status Codes**

HTTP response codes show what happened when a crawler requested a URL.

#### **200 Success**

The server delivered the page successfully. Important indexable pages should generally return this response.

#### **301 or 308 Permanent Redirect**

The URL permanently moved. Internal links, sitemaps, and submission lists should normally use the final destination.

#### **302 or 307 Temporary Redirect**

The redirect is temporary. Persistent use should be reviewed.

#### **404 Not Found**

The requested page does not exist. Broken links and outdated sitemap entries should be corrected.

#### **410 Gone**

The page was intentionally removed and is not expected to return.

#### **5xx Server Error**

The server could not complete the request. Repeated failures may interfere with crawling and reduce website reliability.

Crawl status analysis should focus on patterns. A single temporary error may not create a lasting issue, but widespread or recurring failures require attention.

### **Server Availability and Crawl Performance**

A crawler depends on the website's server to deliver pages and resources reliably.

Analysis should consider:

- Server response times
- Timeouts
- Connection failures
- Error frequency
- Hosting resource limits
- Database performance
- Traffic spikes
- Content delivery configuration
- Regional availability

If Googlebot repeatedly encounters slow responses or server errors, crawl activity may become less efficient.

Improving caching, hosting resources, database queries, and content delivery can support both users and search crawlers.

### **Robots.txt Crawl Analysis**

Robots.txt controls which areas crawlers may access.

A crawl status audit should review:

- Important directories blocked accidentally
- Wildcard rules affecting many URLs
- Development restrictions copied to production
- Critical CSS or JavaScript resources blocked
- Old rules retained after migration
- Different instructions for different bots
- Sitemap declarations
- Conflicts between crawl and index goals

Robots.txt controls crawler access rather than guaranteed index exclusion. If a page is blocked, Google may not be able to process its content or directives.

Pages intended for search visibility should normally remain accessible.

### **Redirect Chain Analysis**

A redirect chain occurs when one URL redirects to another, which redirects again.

Chains often develop after:

- Multiple website migrations
- Domain changes
- HTTP-to-HTTPS transitions
- URL restructuring
- Content management system changes
- Trailing-slash updates
- Uppercase and lowercase corrections

Each extra redirect adds another step for users and crawlers.

Internal links, XML sitemaps, canonical tags, and submission campaigns should point directly to the final preferred URL.

Redirect loops are more serious because neither users nor crawlers can reach the destination. These should be corrected immediately.

### **XML Sitemap Crawl Analysis**

XML sitemaps help crawlers discover preferred website URLs.

A clean sitemap should include:

- Canonical pages
- Successful response URLs
- Useful content
- Current preferred locations
- Pages intended for search visibility

It should exclude:

- Redirects
- Broken pages
- Noindex URLs
- Blocked content
- Duplicate parameters
- Internal search results
- Low-value filters
- Non-canonical versions

Large websites can divide sitemaps by product, category, article, location, image, or video.

Segmentation makes crawl patterns easier to understand. If product sitemaps show more errors than article sitemaps, the product templates may require investigation.

## **Internal Linking and Crawl Discovery**

Internal links guide crawlers through a website and communicate relationships between pages.

Crawl analysis should identify:

- Orphan pages
- Broken internal links
- Links pointing to redirects
- Important pages buried too deeply
- Inconsistent URL formats
- Navigation unavailable to crawlers
- Valuable pages receiving few links
- Excessive links to low-value pages

Important content should receive meaningful links from navigation menus, category hubs, related articles, breadcrumbs, product collections, and contextual sections.

URL submission can support discovery, but it should not replace a logical internal linking system.

### **Crawl Depth Analysis**

Crawl depth measures how many internal link steps separate a page from an important starting location, such as the homepage.

Pages buried deep within the site may be more difficult for crawlers and users to find.

Important pages should generally have a clear route through:

- Main navigation
- Category pages
- Topic hubs
- Breadcrumbs
- Product collections
- Related content
- HTML sitemaps

A shallow structure does not mean placing every link on the homepage. It means creating a logical hierarchy where valuable pages remain accessible through relevant paths.

### **Duplicate URL Crawl Analysis**

Duplicate URL patterns can consume crawler attention and make status data difficult to interpret.

Duplicates may come from:

- Tracking parameters

- Sorting and filtering
- Session IDs
- Print versions
- HTTP and HTTPS variations
- Uppercase and lowercase URLs
- Trailing-slash inconsistencies
- Pagination
- Calendar archives
- Tag combinations

Possible solutions include canonical tags, redirects, parameter management, internal link standardization, sitemap cleanup, and content consolidation.

The objective is to guide crawlers toward preferred pages while preserving useful website functionality.

### **JavaScript Crawl Analysis**

Modern websites often use JavaScript to display content, links, navigation, products, and metadata.

Crawl analysis should verify that:

- Primary content renders correctly
- Important links remain accessible
- Canonical tags are stable
- Robots directives remain accurate
- Mobile rendering works
- Error states do not replace content
- Critical resources are crawlable
- Content does not require user interaction
- Server and client versions agree

A page may look complete in a browser but provide incomplete information to crawlers if scripts fail.

Stable rendering supports reliable discovery and technical interpretation.

### **Server Log Analysis**

Server logs provide direct evidence of crawler requests. They show which URLs were requested, when visits occurred, and what response the server returned.

Log analysis can help answer:

- Which website sections receive crawler attention?

- Are important pages being visited?
- Are duplicate parameters consuming requests?
- Which URLs return repeated errors?
- Are redirects receiving frequent visits?
- Has crawl activity changed after migration?
- Are newly submitted pages receiving requests?
- Which templates respond slowly?

Submission history and server logs measure different events. Submission records show which URLs entered a workflow, while logs show requests that reached the server.

Combining both sources provides a more complete crawl analysis.

### **Crawl Status After Website Migration**

A website migration can change URLs, redirects, sitemaps, internal links, templates, and crawler instructions.

Post-migration analysis should review:

- Old-to-new redirect mappings
- Final destination responses
- Canonical tags
- Robots.txt changes
- Noindex directives
- Updated XML sitemaps
- Internal link destinations
- Missing pages
- Redirect chains
- Server errors
- Priority URL submissions

After validation, important migration URLs can be organized into ExpressIndexing batches. Individual states provide a record of submission activity.

Submission supports discovery but cannot replace accurate redirects and technical migration planning.

### **Crawl Status for Backlink Pages**

External pages containing backlinks can also be included in a crawl monitoring workflow.

Link builders may monitor URLs from:

- Guest posts
- Editorial mentions
- Business directories
- Resource pages
- Partner websites
- Profile pages
- Local citations
- Niche publications

Before submission, confirm that the external page loads publicly, returns a successful response, displays the intended backlink, and offers relevant content.

Indexing does not automatically make a backlink valuable. Source quality, relevance, editorial placement, and context remain essential.

### **Using Historical Crawl Data**

Crawl status changes over time as websites publish, migrate, update, and remove pages.

Historical records can reveal:

- Page types with recurring failures
- Templates producing invalid URLs
- Changes after technical fixes
- Migration recovery patterns
- Differences between new and updated content
- Duplicate submissions
- Campaigns requiring follow-up
- Opportunities for automation

Exportable submission history also helps agencies report which URLs were processed and which links required additional attention.

### **Automating Crawl Status Workflows**

ExpressIndexing provides REST API access for submitting URL batches and retrieving their statuses.

An automated workflow can:

1. Detect new or updated URLs.
2. Validate response codes.
3. Check robots accessibility.
4. Review canonical tags.

5. Remove redirects and duplicates.
6. Confirm sitemap eligibility.
7. Organize approved pages.
8. Submit them through the API.
9. Store individual references.
10. Retrieve updated statuses.
11. Flag unsuccessful requests.

Automation can improve consistency for agencies, publishers, marketplaces, and e-commerce websites.

Broken, blocked, redirected, duplicate, non-canonical, and low-value URLs should be excluded before automated submission.

### **Website Crawl Status Checklist**

Before submitting an important URL, confirm that:

- The URL is correctly formatted
- The page is publicly accessible
- It returns a successful response
- Robots.txt permits crawling
- No unintended noindex directive exists
- The canonical URL is accurate
- The page appears in the correct sitemap
- Internal links point to it
- No redirect chain exists
- Essential content renders correctly
- Mobile access works
- The server is stable
- The content provides original value

After submission, monitor the individual state and investigate failures before resubmitting.

### **Common Crawl Status Analysis Mistakes**

Avoid these common errors:

#### **Confusing Submission with Crawling**

A submitted URL is not guaranteed an immediate Googlebot visit.

#### **Confusing Crawling with Indexing**

A crawled page is not automatically included in Google's index.

### **Repeatedly Submitting Broken URLs**

Technical problems should be corrected first.

### **Relying Only on XML Sitemaps**

Important pages also need meaningful internal links.

### **Ignoring Server Logs**

Logs provide valuable evidence about actual crawler requests.

### **Monitoring Without Corrective Action**

Data is useful only when recurring problems are addressed.

### **Can an Indexing Site Guarantee Google Crawling?**

No third-party indexing site can guarantee when Google will crawl, index, retain, or rank a submitted URL. Google independently controls its crawler and search index.

An indexing platform can facilitate URL submission, display individual states, organize batches, and identify unsuccessful requests. It cannot override technical restrictions or force Googlebot to visit a page.

Reliable crawl performance depends on:

- Accessible pages
- Stable hosting
- Correct response codes
- Accurate redirects
- Clear canonical signals
- Strong internal linking
- Clean XML sitemaps
- Controlled duplicate URLs
- Reliable rendering
- Useful content

### **Final Thoughts**

Website crawl status analysis helps website owners understand how efficiently search engines can discover and access important pages. It identifies submission failures, server errors, crawl restrictions, redirect problems, outdated sitemaps, weak internal links, duplicate URL patterns, and rendering issues.

ExpressIndexing supports the submission and monitoring stage through batches of up to 200 URLs, individual link references, live statuses, exportable records, and API access. These capabilities can

help agencies, publishers, bloggers, developers, e-commerce businesses, affiliate marketers, and link builders manage search discovery campaigns at scale.

The platform works best when combined with technical monitoring, clean sitemaps, strong internal linking, reliable servers, and valuable content. Google remains responsible for deciding when it crawls a page, whether that URL enters the search index, and how it ranks.

By validating pages before submission, monitoring individual outcomes, analyzing server activity, and correcting recurring problems, website owners can create a clearer and more efficient crawl workflow. This structured approach reduces uncertainty and gives important pages a stronger technical opportunity to be discovered and evaluated.